The Educational Effects of Immigrant Children
A Study of the 1998-1999 ECLS-K Survey

MPP Professional Paper

In Partial Fulfillment of the Master of Public Policy Degree Requirements
The Hubert H. Humphrey School of Public Affairs
The University of Minnesota

Hao Jia

May 11, 2015

*Signature below of Paper Supervisor certifies successful completion of oral presentation **and** completion of final written version:*

Judy Temple	4/28/2015	5/11/2015
Typed Name & Title, Paper Supervisor	Date, oral presentation	Date, paper completion

Arthur Rolnick	5/11/2015
Typed Name & Title, Second Committee Member	Date

Signature of Second Committee Member, certifying successful completion of professional paper

Introduction

The United States is the top choice for many potential immigrants because of its booming economy, comprehensive health insurance system, and a reputation for favoring diversity. According to recent statistics, over 38 million people living in the United States were born outside of the country, which equals 12% of the population. In 2001, around 20% children in primary and secondary school had at least one foreign-born parent (Jamieson, Curry, and Martinez, 2001). Moreover, the number of students under 15 years old from immigrant households had increased from 120,000 to 150,000 in the last two decades, and such trend keeps increasing (Homeland Yearbook of Immigrants, 2004-2013). Under such circumstance, it is clear that schools have been becoming more diverse and more interactions between native-born and immigrant children would be observed. As a result, the need for educational policies to provide qualified education for natives and immigrants simultaneously is a key public policy issue.

Previous researchers have suggested that the presence of immigrants has negative effects on natives' quality of education (Kao and Tienda, 1995; Glick and Hohmann-Marriot, 2007, Lopez and Stanton-Salazar, 2001). However, most of these studies were conducted at the high-school level and above, and very few researchers have quantified such negative effects. Considering the negative effects from immigrants, it is reasonable to predict the more immigrant children appear in the class, the worse natives would perform. In order to evaluate the association between natives and immigrants in the elementary school, the key research question this paper addresses is: What is the effects of the immigrant students' share of the student population on the native-born children's academic performance in the elementary school channel? This question is addressed using Elementary Childhood Longitudinal Study, Kindergarten (ECLS-K) dataset from 1998-1999, a national wide survey of children from kindergarten to eighth grade along with their parents and teacher's information. By using such dataset, the immigrants can be easily identified through parents' demographic information. Moreover,

the longitudinal collections of test scores are good indicators for children's academic performance.

This research question is important within the educational framework because the elementary school education will have significant effects on children's life course. According to the "Heckman Curve", the earlier investment in education, the greater the return received by children and society. Such returns are not only limited to good academic performance in the elementary schools, but also include social skills, interpersonal communication as well as teamwork ability (Currie, 2001). Additionally it is a useful indicator for the earning capacity: at a given number of years of education, higher educational quality always corresponds with higher income and better job opportunities (Hunter, 2012).

Better understanding of the association between immigrant and native-born children is crucial to making efficient educational policy and to address the popular concern that a greater immigrant presence is reducing the quality of education in the U.S. American elementary school education exists not in a vacuum but a world that is becoming more globalized. Increased newly-arrived immigrants and their young children generally have limited language proficiency (LLP). If immigrants and natives are taught in the same class, the teachers may slow the pace of instruction to accommodate non-native speakers (Hunter, 2012). Similarly, for immigrant children with unqualified prior education comparing with their native classmates, teachers may lower expectation for all students. (Chin et al, 2012) Under such circumstances, the natives would have lower educational quality if they share the classroom with immigrants. Many people express concern that immigrant children would require too much educational resources, at the expense of native-born children (Fix and Zimmerman, 1993). Thus, increasing the immigrant share of the school would negatively affect natives' education attainment.

Furthermore, some studies argue that such effect might be more significant among low SES native children than those of high SES. Those low SES immigrant families tend

to locate closely to other families of similar SES, immigrant or natives, thus share the similar learning environment. Unlike richer parents, low SES natives would have more difficulties in moving their child to the school with a lower immigrant share or immigrants with better prior education (Betts and Fairlie 2003). Also, low SES parents maybe less able to provide instruction at home in compensation of poor education quality. Therefore, many studies show that native minorities have particular low high-school graduation rates, and such phenomenon is be related to immigration (Orfield et al. 2004, NCES 2008, Noguera et al. 2011).

If above concerns are well founded in the elementary school level, there is a need for reform in the immigrant education to ensure the education quality for natives, especially the minority ones. Reform could include but not limited to provide extra educational resources to high immigrant areas, such as implementing preschool program to improve language skills or strength basic knowledge for elementary schools (Singer, 2008; Garcia, Kleifgen and Falchi, 2008).

For this research paper, I quantify the negative effects of immigrant students on native-born children's academic performance by applying linear regression analysis model. By choosing immigrant share at school level as the main independent variable, I expect that the higher immigrant share would result in worse native's academic performance. Furthermore, in order to detect if native's poor performance is due to demographics characteristics themselves, the propensity score matching is employed to detect if self-selection issue existed.

Data

The ECLS-K dataset was collected by the U.S. Department of Education's National Center for Educational Statistics. It consists of a nationally representative cohort of 21,260 children who entered kindergarten in the fall of 1998 and who are followed until eighth grade. A multistage cluster sampling design with stratification by region (South, Central, North, Northwestern), was adopted for this data set: once the school was selected,

children within that school were selected subsequently. These children were drawn randomly from a nationally representative sample of about 1,000 U.S. public and private schools. This is a long-term survey that covers 7 waves (from kindergarten to 8th grade), which provides a continuous indication of children's academic improvement, as well as parent, teacher and school characteristics.

The detailed table of observations of each wave is provided in the Appendix. Although the questionnaire and interviews were conducted universally, the attrition rate is significant due to the data were collected via a variety of direct and indirect procedures at multiple time points. From the original sample of children sampled in fall kindergarten (wave 1), 5% of them dropped by spring kindergarten (wave 2), and 17% dropped further by the spring first (wave 3). In addition, the information from the parent questionnaire was available for only 80% of the origin sample. Considering in the wave 3, only 30% of the population were sampled, the wave 3 is excluded in this research. According to the user manual, standard missing data codes are used in the ECLS-K data file to identify different types of missing data, for example, a missing data code of -1 indicates that an item was not asked; a -7 missing data code indicates that a participant refused to answer an item, etc. Therefore, this research consists of 14,960 children for whom math and reading test scores were available for all six waves and relevant demographic and school information is available.

Since the ECLS-K dataset was collected from a sample rather than the entire population, not all schools, teachers, and children had an equal probability of selection, nor did all schools, teachers, and children participate in this survey. In order to adjust for differential selection probabilities and reduce bias associated with differential nonresponse, the sampling weight (pweight) is applied for each wave.

Dependent variables: Academic Achievement

For the ECLS-K dataset, each child was given individually administered math and reading assessment at each data collection time point. Because not all children took the exact same test, IRT (Item Response Theory) scores were calculated accordingly, which were collected by asking different questions depending on the answers they provided to the initial questions on the test (Fryer & Levitt, 2004). Math and reading IRT scores are good indicators evaluating academic performance and these scores are comparable across students within a wave and also across waves enabling comparison of children's performance over time. In order to evaluate to what extend, the native-born children's academic performance would be affected by the appearance of immigrants, only the native-born children's math and reading test scores are selected as dependent variable.

Immigrant Country of Origin

Immigrants are identified in the dataset in the following manner. The questionnaire asked the parent respondent in the spring of kindergarten and first grade whether her/his child was born in the U.S. If the parent reported the child was not born in the U.S., then the parent was also asked to report the country of origin. A total of 12 regions were identified in this study based on country of origin. One remarkable point regarding country of origin is that the interview only asked the country of origin of one parent, it is possible that not all children of immigrants would be captured in this dataset, for example, if only native-born mother was interviewed, the child was seen as native-born without knowing if his/her father is foreign-born.

Table 1. Statistics of Immigrant's Region[1]

Country of origin	Observations	Percentage
East Asia (China, Vietnam, India, Indonesia)	184	6.57%
South Asia	376	18.63%
South East Asia	487	1.72%
Africa (Ethiopia, Chad, Sudan)	45	4.63%
North America (Canada)	121	34.12%
Mexico	892	1.80%
Cuba	47	9.14%
Europe (Russia)	239	5.70%
Caribbean	149	2.68%
Puerto Rico	70	7.19%
Central American (Belize, Costa Rica, El Salvador)	188	14.38%
Oceania (Solomon Islands, Marshall Islands)	4	18.63%
Total	2,802	100.00%

Independent variable: Immigrant Status

In order to identify the immigrant children, I follow the existing definition of "1st generation" to include all the children born outside of the United States, and

[1] There are numerous foreign countries of origin from ECLS-K dataset having very few observations. Therefore, I did not include them in Table 1.

second-generation to include all the children born in the United States with at least one parent is immigrant (Glick and Hohmann-Marriot, 2007). The parents' interviewing questionnaire asked the parent respondent in the spring of kindergarten and first grade "If child born in the United States". If the parent reported the child was not born in the U.S., then the parent was also asked to report the country of origin. I group immigrant children into first generation if the answer is yes, and second generation immigrant if the answer is no.

Since the ECLS-K dataset is a survey, only around 20 students are surveyed each school at each data collection time point, it is impossible to calculate the immigrant share at school level by comprising all of student's information. Therefore, to find the generalized estimator of the immigrant share of each school for every wave, I calculate the immigrant percentage among the sampled students:

$$immigrant\ \%_{it} = \frac{no.of\ immigrant\ students_{it}}{total\ no.of\ students\ at\ each\ school_{it}}$$

Where i stands for School ID and t stands for wave number.

Language proficiency

One of the most important criteria used to distinguish native-born and immigrant children is English proficiency. A lack of English proficiency, especially reading and listening skills is mentioned very frequently. In kindergarten and elementary school, most children have not acquired the ability of studying independently and children who lack English proficiency are very likely falling behind teaching process. Therefore, a lack of English proficiency is corresponds with poor academic performance among first- and second generation to a large extent (Rosenthal ,Baker, and Ginsburg, 1983). Therefore, English proficiency might be helpful to capture most discrepancies in test scores results from country of origin as well as generational status.

Corresponding to immigrant percentage, two more school-level variables are employed in this research: the language the immigrant children speak at home, and if immigrant students passed the OLDS test or not. As mentioned earlier, the most significant discrepancies between immigrant and native-born children are language proficiency. With that being said, if native-born child attend a school with high share of immigrant students, it is likely that native-born students would have less educational resource, hence achieve lower test scores.

In the ECLS-K dataset, the language used at home can be found by the question from immigrant parents' questionnaire: Do you speak foreign language to your child at home? Some immigrant parents are able to switch to English-only households for the next generation because they have good education background and good language ability. However, other parents might be unwilling to switch their native language used at home. For those parents speaking foreign language at home, their children do not have a chance to learn English outside the classroom. Therefore, they have more pressure while they studying course material and language in the classroom at the same time. The instructors have to lower their teaching pace in order to ensure immigrant understand the teaching contents.

The other variable is an indicator whether immigrant children passed OLDS test or not. The OLDS (Oral Language Development Scale) test was given to those children who have a non-English language background. The screener determines if children understood English well enough to receive the direct child assessment in English. If children did not pass this test, it is reasonable to believe his/her language skills are not qualified for the course material above the kindergarten.

Both variables can be calculated as the percentage form as similar as immigrant percentage, the equations are:

$$Fail\ OLDS\%_{it} = \frac{no.of\ students\ failed\ OLDS_{it}}{total\ no.of\ students\ at\ each\ school_{it}}$$

$$Otherlanguage\ \%_{it} = \frac{no.of\ immigrant\ students\ use\ other\ language_{it}}{total\ no.of\ students\ at\ each\ school_{it}}$$

Where i stands for School ID and t stands for wave number.

Children's characteristics

Basic characteristics

In order to capture the basic description about children's characteristics, gender, age and race and ethnicity are included. It is very common that American families have more than one child. However, for those immigrant families below poverty threshold, a greater number of children in the family may reduce the resources available to each child. In other words, immigrant child with siblings may receive less parental care and fewer resources.

Beside the number of children, child' gender is crucial to some extent. For those Asia families, many parents prefer boys to girls due to cultural tradition. In that situation, families are going to assign more resources to boys.

Repeating kindergarten

Besides the immigrant children share effect, native-born one's prior education is another crucial factor affecting the academic performance, therefore, repeating kindergarten or not is a good indicator for those children who are not qualified for higher education.

Number of books at home

Books at home are another useful indictor to measure if children will receive good education from their parents. For the analysis, I include a dummy variable for those

children whose families have more than 100 books.

Parents' characteristics

Age:

The first parents' characteristic deserve some attention is age. Young parents are more likely to have very limited income resources, so the living area they can afford might be close to violence, drug and alcohol issues (Alba and Nee, 2003). Especially for 2nd-generation immigrants, parent's age at arrival has significant effect on his/her child's academic performance. Due to the relatively short time staying in the United States, parents are very likely not familiar with education and social systems in the United States. Such a family might not choose the best school for their children. On the other sides, some research showed that parent's more recent arrival will encourage them to spend more time with their children after school to help them adapt to new areas (Kao and Tienda, 1995).

Besides the parents' age, another important index is the parents' income level. Obviously, parents' with low income are not able to afford tuition fee or required study resources. Moreover, poor families might not be able to afford rent for living around a friendly and safe area. From the ECLS-K dataset, families' income was collected and created to a household-level poverty variable. Income was compared to preliminary Census poverty thresholds for 1998, which vary by household size. Households whose income fell below the appropriate threshold were classified as poor (see Appendix Table 2).

Parents' education backgrounds also play an important role in children's early-stage education. Highly educated parents are able to provide instructions to children regarding their homework. To some extend, parents who are familiar with the class content can be "teacher at home," to provide extra guidance beside class-time.

Teacher's characteristics

Multi-language ability:

As mentioned before, if there were a significant amount of student lacking language proficiency in the classroom, it is more difficult for teachers to teach than the homogeneous classroom. Therefore, if a teacher can speak more than English, he or she can switch to other language when explaining complicated concept to immigrant children, make instruction efficient.

Certification:

Teachers who have a certification tend to provide better instructions than teachers have no certification. Some research showed that in mathematics, teachers who have a standard certification have a statistically significant positive impact on student test scores relative to teachers who are not certified in their subject area (Goldhaber & Brewer, 2000). Although such research was conducted in the high-school level, teachers owing certification often were treated as experienced ones, thus would be able to help students perform better.

Descriptive Statistics

The ECLS-K data show that native-born children outperform immigrant peers in Math for all of the waves, and upon the kindergarten, native-born achieve significantly better reading test scores as well ($p<0.005$). Table 2. shows the math and reading test scores for native-born and immigrant children. Table 3. summarizes the descriptive statistics, including weighted and unweighted observations and the total universe for the covariates respectively.

Table 2. Test scores for native-born and immigrant children

Wave	Subject	Native-born	Immigrant	Difference
1	Math	26.885	23.067	3.817***
	Reading	35.522	35.378	0.144
2	Math	37.290	32.621	4.670***
	Reading	46.732	45.784	0.948
3	Math	44.464	38.937	5.527***
	Reading	53.451	51.575	1.876
4	Math	62.237	56.323	5.914***
	Reading	77.733	72.800	4.933***
5	Math	100.341	93.044	7.270***
	Reading	128.875	118.085	10.791***
6	Math	123.497	119.016	4.481***
	Reading	150.774	141.657	9.117***
7	Math	140.955	136.997	3.958***
	Reading	169.832	161.168	8.664***

* $p<0.05$; ** $p<0.01$; *** $p<0.005$

Table 3. Descriptive statistics for native-born children[2]

	Native-born					
	Estimates				Total Universe	
	Mean	SE	Obs	Count	Obs	Count
Father's age	36.287	0.121	13930	12863	17586	19702
Mother's age	33.012	0.105	17241	16293	17586	19702
No. Of Siblings	1.429	0.017	17586	16657	17586	19702
Girls	0.490	0.004	17586	19689	17586	19702
More than 100 books	0.875	0.006	16781	19689	17586	19702
Repeating the kindergarten	0.434	0.015	17586	19702	17586	19702
Mother's education						
Less than high school	0.175	0.007	17586	19702	17586	19702
High school	0.452	0.007	17586	19702	17586	19702
Some college	0.433	0.007	17586	19702	17586	19702
College	0.225	0.007	17586	19702	17586	19702
More than college	0.133	0.006	17586	19702	17586	19702
Low income	0.196	0.010	17586	18515	17586	19702
Private or Public School	0.163	0.009	11792	2273966	14072	2658786
Certificated teacher (wave 1)	0.876	0.008	15123	3097683	16520	3345924
Certificated teacher (wave 2)	0.873	0.009	15330	2908906	16879	3170568
Certificated teacher (wave 4)	0.892	0.009	11457	2517289	13114	2849433
Certificated teacher (wave 5)	0.892	0.010	9381	2336654	10688	2626728
Teachers speak English only	0.921	0.008	15482	2953466	17031	3215129

[2] The method for calculating standard errors is the Taylor Series Linearization Method. This method uses PSU and strata identifiers to calculate the appropriate standard errors.

OLS results

In order to check if there is a association between native-born children's performance and immigrant percentage, an OLS regression is estimated:

$$y_{native-born} = \beta_0 + \beta_1 S_i^{IC} + \beta_2 S_i^{O} + \beta_3 F_i + \beta_4 C_i + \epsilon_i$$

T_i – Native student j[th]'s test score

S_i^{IC}- Immigrant percentage at the i[th] School

S_i^{O}- All other school variables, e.g. Private or Public, location, etc.

C_i- Children's characteristics, e.g. gender, no. of siblings etc.

F_i- Parent's characteristics and education background

By running the OLS regression analysis, I will use a t-test to determine if the high share of immigrant students would associate with low test scores of native-born children. Considering the sample size is decreasing from wave 4 to 7 due to drop out and changing schools, the 0.05 significance level will be employed. Furthermore, I will narrow down the regression analysis to specific racial group: Black, White, Asian and Hispanic, to see if the immigrant share associates with specific racial group children's academic performance.

For the standard errors, the paired-jackknife, or JK2, method is employed as the appropriate replication technique to use. In order to use this method, I need to select replicate weight variables that are associated with the sampling weight variable, and each sampling weight variable in the ECLS-K has a set of replicate weight variables that are associated with it. For example, the replicate weights associated with the sampling weight variable C1CW0 are C1CW1 through C1CW90.

The cross-sectional estimates from spring kindergarten (wave 2), controlling for demographics and other potential covariates listed above, is reported in Table. Immigrant

percentage at the fall semester of kindergarten was not significantly associated with both test scores. Moreover, contradict to the previous hypothesis, for reading test score, the higher share of immigrant had positive effect on native-born children's performance. From the regression outputs, it is not surprising that the association with immigrant share and native-born test scores weakened when we controlled for demographics differences. In our study, having a mother with college diploma is associated with a higher test score for children ($p<0.05$). Therefore, without controlling potential covariates, the association between immigrant share and test scores would be overestimated. However, there may be still unobserved effects between the children enrolled in high immigrant share schools and those not, which may influence their academic performance when they enter primary school. As a result, cross-sectional estimates may still suffer from bias.

Table 4. Cross sectional results for OLS analysis

	Reading	Math
Immigrant percentage	-1.845	1.994
	(1.430)	(1.811)
Father's age	0.0556*	0.0356
	(0.0306)	(0.0368)
Mother's age	0.0793**	0.0808**
	(0.0325)	(0.0372)
No. of siblings	-0.268***	-0.975***
	(0.103)	(0.139)
Girl	-0.505**	2.059***
	(0.240)	(0.270)
More than 100 books	1.594***	2.403***
	(0.384)	(0.368)
Less than High school (reference = college)	-2.814***	-3.202***
	(0.425)	(0.396)
Some college	1.868***	1.941***
	(0.298)	(0.390)
College	2.649***	3.441***
	(0.462)	(0.528)
More then college	-3.485***	-3.125***
	(0.371)	(0.454)
Below poverty threshold	1.824***	2.228***
	(0.550)	(0.751)
Private school	-2.519***	-1.262
	(0.789)	(0.937)
Black (reference=White)	2.368***	0.954
	(0.705)	(0.768)
Asian	3.094***	5.084***
	(1.108)	(1.396)
Hispanic	-0.728	-0.722
	(0.881)	(0.979)
Teacher speaks English only	0.297	-0.680
	(0.581)	(0.857)
Teacher does not hold certification	-0.141	-0.169
	(0.503)	(0.688)
R^2	0.10	0.14
N	10,274	10,289

* $p<0.05$; ** $p<0.01$